EVALUATING
INDIVIDUALISED MATHEMATICS
INSTRUCTION

A thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Master of Arts in Psychology

in the

University of Canterbury

by

M.J. Coppen

University of Canterbury

1980

CONTENTS

ABSTRACT

This study sought to undertake an experimental evaluation of an individualised student paced approach to mathematics learning as compared to the traditional paced method. The effects of mastery and student proctors within the individualised approach were also investigated.

The settin for this investigation were the four classes in the lower half of the third form of a large urban single sex school for boys. One of the major aims was to attempt to replicate an earlier study in a similar setting as well as exploring possible explanations of the previous results. The four equivalent third form classes were all taught by the same teacher. One class was taught in the traditional lecture manner, the other three using variations of an individualised approach. Of the three individualised classes one required no mastery of the material whereas the other two did, one of which was also involved in the use of student proctors.

The individualised method was compared to the traditional on the basis of student achievement on an end-of-year reference test. These results were also used to compare the effects (or lack of) of mastery and student proctors.

It was found that the students in the traditionally taught class performed better than those in the individualised classes. There appeared to be little or no difference in the results between these three classes.

When the results for the four classes were re-analysed for exposure to the course material which comprised the test, it was found that all classes performed similarly.

From the results of this investigation it is argued that the major factor in student achievement is exposure to material and not the mode of instruction, or in the case of the individualised classes the added variables of mastery or student proctors.

In conclusion, the argument is advanced that evaluating a teaching method by comparing student academic outcomes only will result in little useful information being obtained about the merit or otherwise of that method. An alternative approach to evaluating a teaching method is presented and its application to the setting of this investigation outlined.

ACKNOWLEDGEMENTS

The author would like to express his appreciation to Neville Blamplied for his supervision of this thesis.

I would like to thank Kevin Quinn for organising the Christchurch Boys' High School timetable such that I was able to carry out this investigation.

My thanks also to Barbara Greenfield for her help with the writing of this thesis.

Lastly but not least, I would like to thank Maryrose Heffernan for deciphering the drafts and typing this thesis.

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1. General Introduction

The teacher of third form mathematics classes in New Zealand is usually required to instruct groups of 25 to 35 students. Very commonly the class is taught as a single group. Where the class is being taught as a group the typical mathematics period begins with a short 'lecture' or worked example and this is followed by assigned problems and homework, with the teacher providing help for individual learning difficulties. When the teacher feels the time is appropriate, a test covering the material being taught is administered.

This type of group approach makes little provision for differences in mathematical skills, background and motivation. Consequently, for some students the information presented is beyond their level of comprehension. This situation may give rise to the instructional problems of inadequate preparation for later topics. For example, many branches of mathematics are sequential in nature, thus a student who fails to master skills required early in a sequence may be quite unable to master skills required later in the sequence. In addition, there is the possible problem that where the information being taught is redundant to a student or when the student's background is such that he/she is sorely deficient in the topic being taught, behavioural problems may arise.

One approach to minimising the problems of differing student skills, background and motivation is to attempt grouping of students in 'streamed' classes. However, streaming is often inadequate in overcoming those instructional problems due in part to the difficulty of producing a homogeneous class along such dimensions as motivation, mathematical aptitude and background.

In some groups of the teaching profession there has been a call for a move away from rigid streaming practices to either total non-streaming or variants of partial non-streaming, for example grouping classes in two or more broad bands of ability. Some authors have called for this change on social grounds, claiming 'less racial segregation, less tension because of lack of emphasis on competition and many social advantages' (John, 1977). However, in moving to non-streamed classes with their attendant wide ranging ability levels, backgrounds and motivations, the instructional problems previously mentioned are often aggravated.

Scriven (1975) has argued that instructional problems in the non-streamed classroom may be overcome by schools individualising their classroom programmes. Scriven sees individualisation as alleviating the pedagogical problems of having respect for individual differences in learning rates and development, and also improving some political problems such as attempting to salvage an adequate education for those students of average and above average ability, when grouping on ability produces racial segregation.

Scriven also argues that, 'the pendulum of education swings to and fro (sic) amongst inovations, and to get away from the worship of the pendulum we must identify and demonstrate merit (or its lack) in such innovations'. That is, we need to evaluate individualised programmes.

The strategy usually employed in the evaluation of innovative programmes is what Scriven (1973) has termed the 'pay off' approach. That is, a comparison of two or more groups, of students outcome measures such as, achievement, attitudes to the subject (or type of instruction) and retention of learning.

To date there have been a large number of such studies in the area of individualised mathematics instruction. Two review articles, Miller (1976) and Schoen (1976), have brought together many of these studies carried out (almost exclusively) in the U.S.A.

Miller (1976) reviewed one hundred and twenty-six studies comparing individualised and traditional mathematics programmes at the primary (N=88) and secondary (N=33) and mixed (N=5) levels. Unfortunately he did not distinguish between the settings. The results of this review are generally ambivalent (see Table 1).

TABLE 1

Comparison of
Individualised and Traditional
Mathematics Instruction
(Miller 1976)

Variable Compared	Superiority for (Percentage) Individual-isation	Traditional	N.S.D.*1
Achievement	36	16	48
Attitude	21	3	76
Retention	20	0	80

*1 No statistically significant difference between the groups at 0.05 level

The only consistent feature he reports is that of a steady decline in comparative achievement results as the duration of the study increased.

Schoen (1976), in his review of twenty studies, dealt only with secondary (N=12) and post-secondary (N=8) settings, distinguishing between the two. In his choice of studies Schoen was more stringent than Miller, in determining both what constituted individualised instruction (see page 14of this thesis) and what constituted 'vigorous' experimental design (see page 15 of this thesis). Although Schoen reviewed fewer studies than Miller he came to similar conclusions (see Table 2). If anything, because of the more stringent criteria he employed when deciding to include a study in his review, his conclusions may be seen as more dismal than Miller's.

TABLE 2

Comparison of
Individualised and Traditional
Mathematics Instruction
(Schoen 1976)

Variable Compared	Superiority for (Percentage) Individualisation	Traditional	N.S.D.*1
Achievement	8	25	67
Attitude	0	25	75

*1 No statistically significant difference between the two groups at the 0.05 level.

Schoen pointed out that there is the possibility that a number of studies showing no significant difference may have actually favoured the traditional group had it not been for the possible intervention of the Hawthorn (or novelty) effect (Neale and Libert, 1973). He also commented that most of the studies reviewed were carried out by enthusiasts for individualisation, thereby raising the possibility of experimenter bias (Rosenthal,1966).

When considering the results of these wide ranging reviews as to the worth of individualisation, a confounding problem is the criteria that the reviewers used for inclusion of studies in their review. It could be argued that if the criteria are sufficiently loose then any possible effects may be rendered unobservable by the vagueness of the experimental design.

It would appear that if we go back to the origins of the comparative approach, the agricultural model, one of the key criteria is to be able to define objectively the fertiliser (treatment) !

The problem is how stringent or wide ranging were Miller and Schoen in accepting definitions of individualised (and traditional ?) instruction for purposes of their reviews ?

Miller (1976) chose as his definitions of individualised and traditional instruction the following (after Crosby 1970):

> 'Individualised instruction is defined as that in which each student participates in setting his own goals, works at his own rate (either alone or as a member of a small group) and participates in evaluating his own progress. Traditional instruction is defined as all methods in which the students are taught as a class. It includes homogeneous or heterogeneous grouping and does not preclude the use of audio-visual aids, committee work or any other techniques traditionally used by the teacher to help students learn.'

White (1972) defined individualised mathematics as:

> 'Not being synonymous with independent study. Individualised instruction means that the student has been matched to an instructional system such that he is working at his own speed, learning style and ability level on appropriate materials in keeping with his goals supported by adequate assistance in a suitable learning environment'.

Schoen (1976) defined an individualised programme as exhibiting the following features:

> 'First, they would be based on a specific set of behavioural objectives. Second, the mathematics content would be divided into small modules or units. Third, learning packets would be written for each unit, the learning packets serving as guides for the students, enabling them to proceed more or less independently through the content at their own pace. Fourth, for the most part the students learn independently from text books and work sheets, though

> some programmes included other media. Fifth, each packet contained pre-tests and post-tests, the student being required to pass one or both before proceeding to the next unit.

From these three definitions it is quite clear that individualised instruction in mathematics is not a simple concept, and in fact means different things to different people. This problem of definition raises an element of uncertainty when lumping individualised studies together for review purposes.

Another problem when considering the results of such reviews is that of poor experimental design. In particular there is the problem of non-randomisation. Miller did not indicate what criteria with respect to the experimental design he used when deciding to include a study in his review. As has been argued 'randomly assigning subjects reduces the likelihood that differences among the groups after treatment will be due to initial differences in the samples rather than to true experimental effects' (Neale and Liebert, 1973). Thus the question arises, of the studies in Miller's review, how many of the results are due to initial differences in the samples?

Unfortunately Miller gives no information about the research design criteria he employed when selecting studies for his review, hence there is no way of evaluating the seriousness or otherwise of this objection to his review.

Schoen however, was apparently more stringent in that he included only those studies which incorporated random assignment of students or statistical equivalence of groups (e.g. by such techniques as Analysis of Covariance ANCOVA). In relation to the latter technique (ANCOVA) there is the controversy of just how appropriate it is (e.g. Evans and Anastasio, 1968), and as Schoen gave no information as to the conditions when ANCOVA was applied or to which studies, there remains a degree of uncertainty about how to interpret his findings.

Another serious problem for the comparative strategy is that of content. Was the content taught to the two groups the same and were the differing topics given equal emphasis in the two situations ? Obviously comparing two groups taught differing contents or having differing degrees of topic emphasis will raise the question of just what was being compared. A related problem is that of testing. Were the groups compared on a criterion referenced test or on a standardised test ? For a criterion referenced test, difficulties in interpreting the results arise if the respective course contents and/or their respective topic emphasis has differed. If, as Walker and Schaffarzick (1974) argue, test results reflect what was taught, then bias towards that group whose course most closely reflects the test would result. The use of a standardised test immediately raises the objection that the two groups are not being explicitly compared on what was taught, thus something other than that method of instruction is being compared. If a standardised test was used, then would any difference between the two groups be expected ?

A problem linked to differing contents is that of differing teachers. Although Schoen (1976) claimed that most of the studies he reviewed were carried out by enthusiasts for individualisation he did not identify which ones. It might also be asked, are the two groups of teachers as committed to the two approaches ? Had they similar training and experience and enthusiasm for the two methods ? However, it can also be argued that teacher variability is a 'straw person'. Stephens (1976) has argued that although two methods may be 'different' what the teachers do may be the same. For example, although one class is using an individualised approach and the other being taught in the traditional manner it is most probable that BOTH sets of teachers will accentuate the important aspects of the curricula to be learnt (and later tested and used as a comparative measure of the merit of the two programmes). As well, Stephens feels they will both positively reinforce learning by approval when it takes place. In other words, the basic human interactions will be the same in both types of classrooms. This again raises the question of just how different the two groups of classrooms are, and just what this difference means to academic learning.

A problem related to the above areas of content and teaching is the quality of instruction. Although it might be considered a 'straw person' it raises the problem of time actually spent on learning. Consider the case of a teacher of an individualised class. Does the study guide contain an equivalent amount of explanation, and the equivalent worked examples, exercises and tests as the teacher would have employed in the traditional approach ? Is the teacher spending more time answering questions about the study guide than he would about the same topic if he had just delivered the traditional lecture ? A related question is, what are the students doing ? For example, is the time the students spend on-task the same in the two settings ? In the individualised classroom are the students spending less time on-task, waiting for explanation or help with the study guide or waiting for their tests to be marked ? Conversely in the traditional class, are the students spending less time actively involved in working problems owing to their sitting passively and listening to the lecture ? Whilst this is only a superficial treatment of the topic of quality of instruction it may raise serious problems for the comparative approach, particularly if, as has been claimed, (Carroll, 1973) time on-task is a crucial variable in student learning.

In summary, the major criticisms of Miller's and Schoen's reviews are the lack of control for the following:

(i) Clear definition of research design. In Miller's review was randomisation involved in assigning subjects to the two modes of instruction ? For Schoen's review there is the problem of whether or not conditions were appropriate for ANCOVA to be used.

(ii) Criteria for inclusion of research studies. In Miller's case were the criteria for individualisation so loose and wide-ranging as to make it meaningless to talk of individualised instruction ?

(iii) Academic content. Within each study was the academic content the same for the experimental and control groups and did it have the same relative emphasis ?

(iv) Comparative measures. Within each study were the experimental and control groups compared on criterion referenced or standardised tests ?

(v) Were there possible confounding teacher differences ?

(vi) Was the quality of instruction the same in the individualised and traditional classrooms?

2. Nature and Scope of the Investigation.

The present study was designed to attempt to explore and where possible overcome the objections noted above in the critque of comparative studies, as well as to attempt a replication of an earlier study in a similar setting (see Coppen, 1976, Experiment 2). The earlier study involved a third form setting in a large urban single sex school for boys and compared student achievement on an end-of-year criterion referenced test (one class using the individualised mastery instruction (I.M.I.) approach, the other three acting as controls).

One criticism of the review studies was the looseness of the criteria by which a procedure was deemed as being individualised.

For the purposes of this thesis the traditional and individualised procedures are outlined below, as also for the previous study (Coppen, 1976). The traditional class is very commonly taught as a single group. Where the class is being taught as a group a typical sequence of events is given in Figure 1. Typically the mathematics period begins with a short 'lecture' or worked example and this is followed by assigned problems and homework, with the teacher providing help for individual learning difficulties. When the teacher feels the time is appropriate, a test covering the material taught is administered.

With I.M.I. the sequence of events is that outlined in Figure 2. Within any given unit the student studies the prepared notes and does the assigned problems with the teacher providing help for individual learning difficulties.

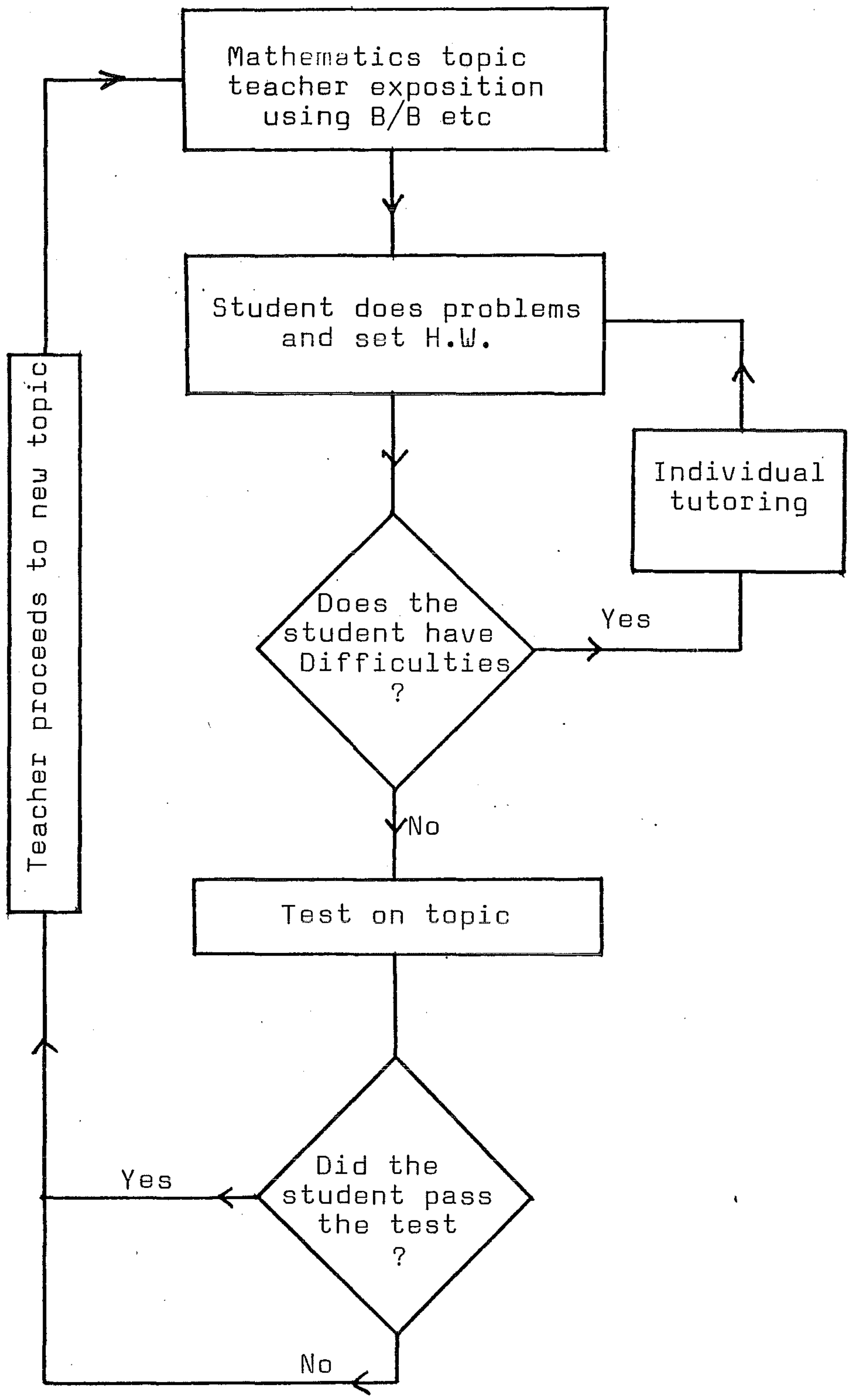


Figure 1
Learning as a class

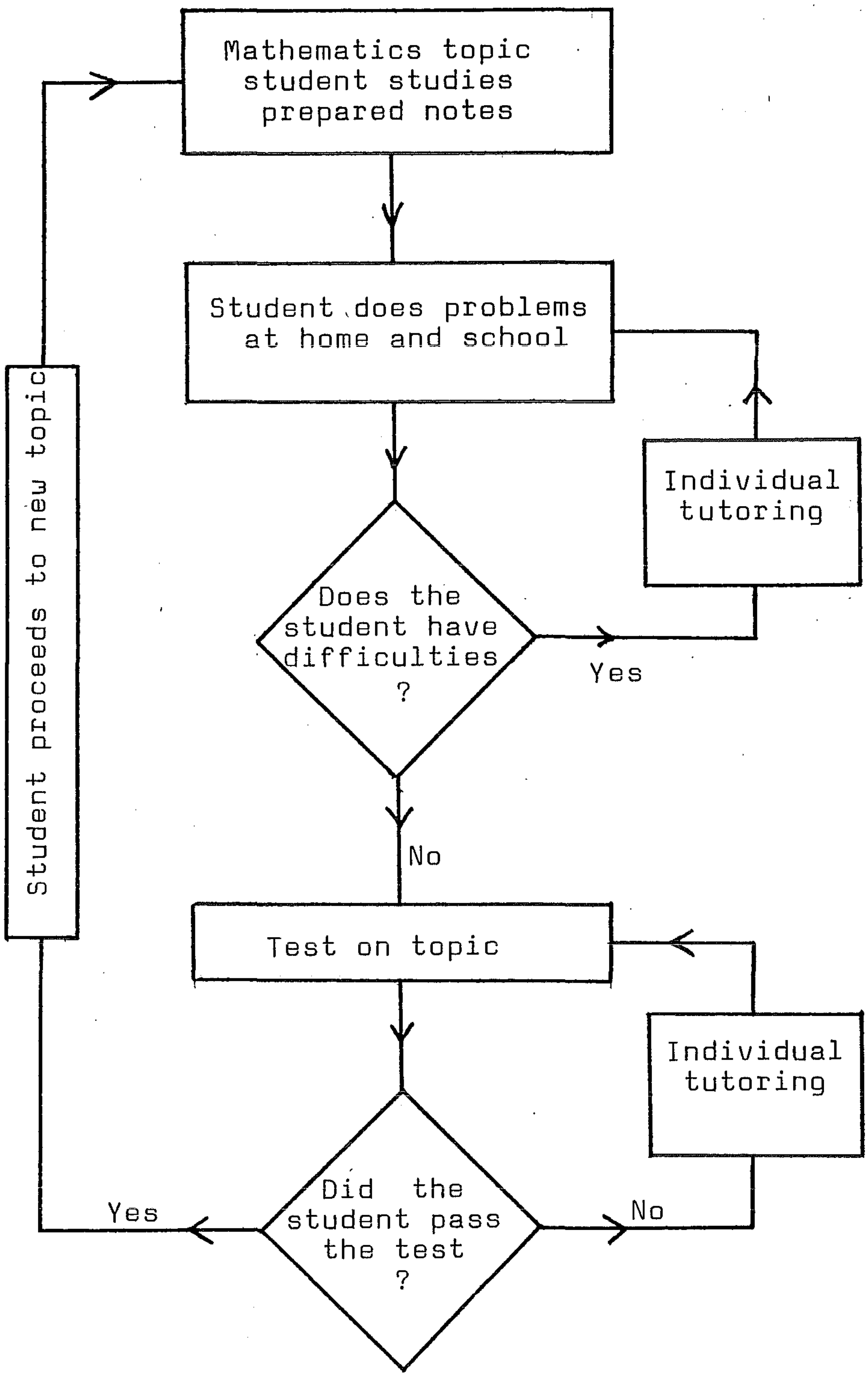


Figure 2
Learning as an Individual

When the student has covered the appropriate work for a topic, believes the work has been mastered, and can provide evidence that the work has been marked and where necessary corrected, he obtains a mastery test from the teacher. If he passes this test he proceeds to the next topic, or if it was an exit test, to the next unit. If the student fails the mastery test, individual tuition and remedial work is provided by the teacher. If the student fails an exit test, remedial work is provided and another exit test is administered. Since each student is free to work at his or her own pace, students may be at different points in the programme. During the maths period a student may be doing any number of things. He or she may be reading the notes of a new unit of work, doing the assigned problems, obtaining a test from the teacher, or waiting for the teacher to clarify some point in the notes or to assist with a problem.

The essential difference between the typical pattern of group instruction and I.M.I. are as follows. With the group instruction the time available for individual help by the teacher is reduced by the teacher's 'lecture'. With the I.M.I. class the total class time is available for individual tuition. The problems of information redundancy for the student taught by the group instructional method no longer exist in the I.M.I. approach, as the more able students quickly pass on to new material. Information deficiencies which arise for the less able students in the group approach and which often result in difficulties with the course content can be readily overcome in I.M.I. either by individual tutoring or by the use of remedial programmes. Generally, the mode of instruction for the group approach is the blackboard or overhead projector. It is not uncommon for the student of low ability to take rather poor notes, that is, notes which are of little use for later reference. This problem obviously does not occur with I.M.I. With the group approach all the students are tested at the same time. As a consequence some students are required to sit tests before they have mastered the skills being taught. With I.M.I. it is the student who decides when to sit each test, provided there is evidence of mastery of work to be tested. Finally with the group approach, students who perform poorly on a test, typically proceed with the class to the new unit of work, rather than being provided with appropriate remedial instruction. In I.M.I. those who fail a test have their problems diagnosed and appropriate remedial work is provided.

The method of assigning students to a class in the 1976 study, Experiment 2, was to rank order the students on the combined results of English and Mathematics achievement tests and to split them into two bands of three and four classes respectively. Within the lowest band of four classes the student ranks were grouped in fours and then from each group of four a student was randomly assigned to a class.

The results of this experiment are shown in Table 3 and suggest that an individualised mastery programme results in significantly higher levels of student achievement on a criterion referenced test than the group instructional procedure used in the control group classes.

TABLE 3

The means and standard deviations for the four classes involved in the Comparison of Individualised and Traditional Mathematics Instruction (Coppen, 1976)

Instructional Method	N	$\bar{X}$(%)	S.D.
Individualisation	35	67.9	18.0
Traditional	36	56.7	13.2
Traditional	35	55.7	11.4
Traditional	36	51.8	14.6

However, it should be noted that this study failed to overcome a number of design problems outlined above:- differing academic content, testing experiences and teachers.

Although all teachers taught to a common syllabus it is impossible to know if the differing topics in the course were given similar emphasis in each of the classes. Unfortunately an interclass comparison for each item on the first comparison test was not carried out. The potential problem of differing content emphasis is aggravated as the criterion referenced test used to compare classes was composed by the teacher of the experimental class (the other three teachers concurring with its content). If, as has been argued previously on page 16 of this thesis patterns of student achievement reflect patterns in the curricula (Walker and Schaffarzick 1974) then a serious objection to the previous results found by Coppen is that the results simply reflect differing contents and not differing modes of instruction.

An allied problem is that of item format construction. The format of the questions (language, style of answer etc) used in the criterion referenced test were similar to that used in the topic tests for the I.M.I. course (often the items in the criterion reference test used the same words, diagrams etc as in the I.M.I. topic tests, only letters and/or numerals being changed). This again raised the possibility that the results Coppen found may be partially explained as a result of prior testing experience and not due to differing modes of instruction as claimed.

Another possible problem arises from possible differential teacher enthusiasm and commitment to teaching at third form level. The control class was taught by a fourth year male teacher. The control classes were taught by male teachers of differing experience. One was taught by a first year teacher. The second was taught for the first term by a retired secondary school principal who had also been a senior mathematics teacher and the remaining two terms by an experienced eighth year teacher. The last class was taught by a teacher with more than ten years experience. Of these classes three (including that of the experimental class) were taught by full-time mathematics teachers; the fourth teacher (first year) taught mathematics to this form only. It may be argued that as the full-time mathematics teachers taught classes ranging from the senior to the junior school that third form classes might not have had a high priority with regard to a teachers total commitment. As the teacher of the experimental class was an enthusiast for individualisation the possibility arises that the results may simply reflect differential teaching enthusiasm and commitment to this level of schooling.

Another possible interpretation of the results is that they are due simply to sampling errors. The Progressive Achievement Test (Mathematics) (P.A.T.) results were not available for all four classes so it was not possible to investigate this objective.

The study reported in this thesis sought to control for the varables mentioned in relation to the previous study (Coppen, 1976), by using the same content (including tests) and the same teacher for all classes included in the evaluation. The possibility of sampling error was investigated by administrating P.A.T. and a pre-test (the 1977 end-of-year criterion reference test)at the beginning of the year.

In addition, this study sought to investigate some other aspects of individualised instruction: mastery and student proctors. One of the underlying principles of the I.M.I. approach is that of mastery, a concept of inherent appeal in a skill orientated subject such as mathematics. A basic problem arises with the concept of mastery, which is if the students attain mastery on 'unit' tests do they retain this mastery over time ?

The answer to the question of whether or not mastery was retained over time was sought by using items of similar format (but differing content) in both the mastery tests at the end of each unit and the end-of-year criterion reference test. If mastery was retained then it might be expected that student total scores (for the mastery classes) on the criterion referenced test (adjusted for those units which the student had mastered) would be equal or exceed the mastery criterion of 80%.

Often teachers using the individualised approach with its frequent testing report a problem of marking tests and being available for tutoring individuals. One way around this problem is the use of so called student proctors, that is to use the advanced students to mark other less advanced student's tests and to tutor them on points of difficulty with the tests.

Two issues arise in the use of students' peers as proctors. Firstly, what is the effect on the students tutored and secondly what is the effect on the student proctors ?

Numerous studies have shown positive gains by students tutored by older students. Cloward (1967) reported a study where fourth and fifth graders were tutored on reading skills by tenth and eleventh grade students. The tutored students improved their reading skills more than a comparison group of untutored students. Johnson and Bailey (1974) used fifth grade students to tutor kindergarten children in arithmetic. They found that the tutored group made far greater gains on a skill-based arithmetic test than did a non-tutored comparison group.

A second group of studies has shown positive gains by students when they were tutored by their classroom peers. Harris and Sherman (1973) using a peer-tutoring situation in fourth and fifth grade mathematics classes found students who were tutored by their peers performed better than when in an untutored situation. Dineen, Clark and Risley (1977) studied the effects of peer-tutoring on the acquisition of spelling words for nine to ten year old students in an ungraded open plan environment. They found a substantial improvement in the number of words spelt correctly by <u>both</u> the tutor and tutee when compared to acquisition of spelling in a non-tutored situation.

This finding of Dineen, Clark and Risley (1977) of improvement in tutors performance is directly related to the second issue, what effect does tutoring have on the student proctor ? If time spent on learning is a crucial variable (Carroll, 1973) then it may be objected that using students as proctors would inhibit their academic achievement through decreasing their time spent on the subject. One answer to this objection is that the revision and 'insight' gained by the student proctor from the situation will more than compensate for loss of time spent on learning new material. Using the more advanced students as tutors ensures that these students are tutoring across the entire course that they have completed, thus to a degree they are in a state of constant revision. Allied to this is the 'insight' gained by having to be able to explain the processes used to other students. A common experience of teachers is that it was only when they had to explain something to others (e.g. a person or class) that they felt they had understood it themselves. A second source of rebuttal comes from empirical studies, for example the previously mentioned study by Dineen et al.(I977). Another study showing positive gains for both tutor and tutee was that of Hamblin, Hathaway and Wodarski (1971). They found in their study of peer tutoring in a grade five mathematics classroom that both the tutor and the tutee benefitted in comparison to a class where there was no peer tutoring.

CHAPTER II

METHOD

Setting:

This experiment was undertaken in a large urban school, Christchurch Boys' High School. In 1977, the year in which the experiment was undertaken, the third form was divided into two broad bands on the basis of English and Mathematics achievement tests and the Otis Intermediate test of general ability, both bands consisting of four classes. The students were quasi-randomly assigned to a class within a band by the method previously outlined (see page 24 this thesis).

Subjects:

The subjects in this experiment were the four bottom band classes. Although the third form had been split into two bands, the classes within the bottom band still exhibited extreme ranges in mathematical ability as measured by the Progressive Achievement Test (Mathematics). For example, P.A.T. percentile scores of the students in the experimental class ranged from 0 to 80 or more.

Materials and Procedures:

At the beginning of the year the P.A.T. and a pre-test (the 1977 end-of-year criterion reference test) were administered to the four experimental classes. The choice of instructional mode was randomly decided. All four classes were taught by the same teacher.

Classes A,B, and D were taught using the I.M.I. procedure and Class C was taught using conventional group instructional procedures (on pages 20 to 24 of this thesis.

The four bottom band classes used the same basic text: Mathematics: A Study in Pattern 1. (Nightingale, Cornwell, Kibblewhite, Gray and Smith, 1972). The course consisted of twelve units based on nine chapters of the text and was supplemented (as were the other twelve units at level 3 - see Figure 3) by teacher-designed units.

The I.M.I. programme consisted of three levels for most units (see Figure 3) and was arranged in five blocks. Within each block some students worked through the level one units for that block. Those who completed this level before the block one test (this test being common to all the classes in both bands and consisted primarily of level two and three material) then returned to the initial unit of that block at level two and worked through it. Other students who quickly and accurately completed the first level one unit (in the first block) and for whom much of this level was simply revision, used subsequent level one units for revision only. They worked on the level two material within block one.

Students who completed the level two material in block one before the block one test then returned to the initial level three unit (unit four) in block one. They then worked through the level three units. Those who managed to complete all the level three units within block one prior to the block test moved onto the level two material of block two. A student who was working at level three but did not complete block one prior to the test simply left that unit incomplete and moved onto unit six in block two at level two after the block test.

Students who were working at level two at the time of the block one test, but who had not completed unit five and who had previously completed the unit five at level one did not complete the block at level two but proceeded directly onto unit six in block two. Those who had quickly and accurately completed the first level one unit and who had been promoted directly to level two work but who had not completed unit five at level two at the time of the block test subsequently dropped back to level one and completed those units in level one in block one before proceeding to block two at level one.

The way a student worked through the rest of the programme was similar to that for block one.

A student who started block two at level two could at the time of the block test be: still working at that level, in which case he would drop down to level one and complete the block at that level and then continue onto the next block at level one; have started level three work and not completed all units, in which case he would not complete that unit but proceed to the first unit of the next block at level two; or he may have completed the level three work in which case he could either be spending his time revising for the block test or working on level two units in the next block.

A student who started block two at level one could at the time of the block test be: still working at level one in which case he continued to work on that unit and then the next unit in sequence in block two; have completed the level one units and be working (but had not completed) the level two units in which case he would not complete that unit but proceed to the first level one unit of the next block; or he may have even reached level three work in which case he would proceed as if he had started the block at level two.

Students who had not started block two level one at the time of the block two test continued to work their way through the course sequentially at level one. For those students special individualised tests were prepared involving the material they had worked through and were given to them at the time of the block tests.

	Unit	Description	Level 1	1	3
Block 1	1	Angles	✓	✓	
	2	Triangles	✓	✓	
	3	Parallel Lines	✓	✓	✓
	4	Polygons	✓	✓	✓
	5	Construction	✓	✓	✓

Block 1 Test

End of Term 1

Block 2	6	Congruency	✓	✓	✓
	7	Statistics	✓	✓	✓
	8	Probability	✓	✓	✓

Block 2 Test

Middle to Term 2

Block 3	9	Sets	✓	✓	✓
	10	Operation with W,I	✓	✓	✓

Block 3 Test

End of Term 2

Block 4	11	Relations and Graphs	✓	✓	✓
	12	Fractions and Decimals	✓	✓	✓
	13	Mensuration	✓	✓	✓

Block 4 Test

Middle of Term 3

Block 5	14	Sentences and Equations	✓	✓	✓
	15	Number bases	✓	✓	

End of year criterion reference test. End of Term 3

Figure 3 Course outline for 1977 bottom band third form classes. (A tick indicates that a unit was written for this level)

The 'traditional' programme utilised the same content, worked examples, exercises (the I.M.I. programme notes being used verbatim in the 'lecture' segment of the sequence) and tests. The only difference between the two approaches was in the sequencing of the levels. In the I.M.I. approach the student worked through a sequence of units at the same level within a block whereas in the traditional programme the levels were treated successively for each unit (level one being followed by level two and then by level three before moving onto the next unit). However, due to time restrictions, for units eleven to fourteen those students who displayed 'mastery' at level two on a class test completed the level three material using the I.M.I. programmes. These programmes were completed by the particular students out of school time. It should be noted at this point that the end-of-year criterion reference test contained <u>only material from levels one and two</u>.

Classes A and D followed the normal programme requiring 'mastery' to proceed from one unit to the next. For Class A at the end of block one the six students farthest through the course (who incidentally had negligible absenteeism records) became the student 'proctors'. Student proctors tutored one period per week. This tutoring involved marking other students' tests (on units the proctors had mastered, marking from a master marking schedule). For those tests with only minimal errors the tutor attempted to elucidate how the student solved the particular item, outlined where the student made his error and showed him the correct method. The student then corrected the item and continued onto the next unit assuming he had 'mastered' the test, otherwise he was required to take another test. Those students with serious problems were referred by the student proctor directly to the teacher. However, it ought to be noted that to obtain a unit test the student in all the I.M.I. classes had to show the teacher his exercise book with the completed problems and a self constructed and completed check test (using the worked examples from the programme notes as questions).

At the beginning of the year, in the second week of the first term a common criterion test (the same test that was to be administered at the end-of-year, see Appendix) was administered under examination conditions to the four classes on the same afternoon. Those students who were absent were not tested later on this pre-test instrument due to problems of arranging a suitable time (the test was of two hours' duration) and possible contamination from discussions with classmates. Later in the first term, in the fifth and sixth weeks, the Progressive Achievement Test (Mathematics)was administered to the four classes. Students who were absent at the time were subsequently tested the following week. (The P.A.T. was of 40 minutes duration and its administration could easily be accommodated in the school programme, unlike the pre-test).

At the end of the year a common criterion exam was administered to the four classes as their final exam (note, this had been earlier used at the beginning of the year as a pre-test).

CHAPTER III

RESULTS

At the beginning of the year there were 127 students in the four lower band classes. Two of the students were administered neither the P.A.T. nor the pre-test. One left the school permanently in the first two weeks of the school year; the other was continually absent for the first half of the year, so it was impossible to arrange times for him to be administered either test. One student transferred from another school in the second term into the experimental setting. Of the remaining 125 students two left permanently in the second term, another was absent for six months on an overseas trip and two others were absent for the end-of-year criterion reference test (hereinafter referred to as the post-test). These students' P.A.T. scores only have been included in analysis. For the pre-test five students were absent from school on the day it was administered and have therefore been excluded from the pre-test analysis. As they were present for the post-test they have been included in the appropriate analyses.

The mean scores on the P.A.T. for the 125 students who began the course are given in Table 4. There were no significant differences in the distributions of the P.A.T. scores for the four classes, $F(3, 121) = .61$, $p > .05$.

TABLE 4

The means and standard deviations of the Progressive Achievement (Mathematics) Test scores for the students initially assigned to classes.			
Class	N	$\bar{X}$(%)	S.D.
A	31	39.4	15.1
B	32	43.5	12.4
C	32	39.3	14.8
D	30	41.3	15.4

The mean scores on the P.A.T. for the 120 students who completed the course (including the post-test) are given in Table 5. Again there was no significant differences in the distributions of the P.A.T. scores for the four classes $F(3,116) = .84$, $p > .05$

TABLE 5

The means and standard deviations of the P.A.T. scores for the students who completed the course.			
Class	N	$\bar{X}$(%)	S.D.
A	29	39.6	15.2
B	32	43.6	12.3
C	30	38.2	14.6
D	29	41.7	15.5

The mean scores on the pre-test for the 115 students who completed the course (including the post-test) are given in Table 6. As with the P.A.T. scores for this group there was no significant differences in the distributions of scores for the four classes $F(3,111) = .51$, $p > .05$

TABLE 6

The means and standard deviations of the pre-test scores for the students who completed the course.

Class	N	$\bar{X}$(%)	S.D.
A	28	18.0	12.0
B	30	21.1	10.6
C	30	19.6	9.1
D	27	20.4	7.8

The results of the P.A.T. for the initial classes suggest that the assignment of students resulted in homogeneous groupings. Also the results of the P.A.T's pre-tests for the course survivors suggest that those students remaining also constituted homogeneous groupings.

The mean scores of the four classes on the post-test are given in Table 7. There was no significant difference in the distribution of the four classes $F(3,116) = 1.45$, $p > 0.05$. However, from Table 7 it can be seen that the traditionally taught class had a higher mean mark than did any of the control group classes, whereas it had the lowest mean P.A.T. mark for those students who completed the course (Table 5).

TABLE 7

The means and standard deviations o the post-test for the four classes.			
Class	N	$\bar{X}$(%)	S.D.
A	29	57.5	23.6
B	32	60.4	22.0
C*1	30	67.0	21.7
D	29	56.0	20.9
*1 Class C was the traditionally taught class.			

The post-test results for the 120 students who completed all aspects of the course were subjected to an analysis of covariance with the P.A.T. scores. The adjusted means are given in Table 8. There is a statistically significant difference in the adjusted distributions of the four classes $F(3,115) = 4.64$, $p < .05$

TABLE 8

The post-test means of the four classes adjusted for covariance with the P.A.T.		
Class	N	$\bar{X}$(%)
A	29	58.7
B	32	57.6
C*1	30	69.8
D	29	55.0
*1 See Table 7		

For the 115 students who completed both the pre- and post-tests, the post-test scores were subjected to an analysis of covariance with the pre-test scores. The adjusted means are given in Table 9. Again there was a statistically significant difference in the adjusted distributions of the four classes $F(3,110) = 3.4$, $p < .05$

TABLE 9

The post-test means of the four classes adjusted for covariance with the pre-test.

Class	N	$\bar{X}$(%)
A	28	60.2
B	30	58.1
C*1	30	67.3
D	27	55.4

*1 See Table 7

P.A.T. and post-test scores for the students in the four classes correlated hightly $p\ (118) = .64$, $p < .05$ as did their pre- and post-test scores $p(113) = .72$, $p < .05$. These correlations indicate that both the P.A.T. and pre-test are good predictors for the post-test scores.

A comparison of the mean scores for the various tests is given in Table 10. The change in ranking of the class means from the P.A.T. and pre-tests to the post-test for the traditionally taught class agree with the results from the analyses of covariance.

TABLE 10

The summary of the means and the rank of the means for the four classes on the various tests.

Class	TEST P.A.T. (5)[2]	Pre (6)	Post (7)	Post with P.A.T. (8)	Post with Pre (9)
A	39.6,3	18.0,4	57.5,2	58.7,2	60.2,2
B	43.6,1	21.1,1	60.4,3	57.6,3	58.2,3
C	38.2,4	19.6,3	67.0,1	69.8,1	67.3,1
D	41.7,2	20.4,2	56.0,4	55.0,4	55.4,4

*2 The figure in brackets refers to the table from which the means are taken.
*3 The first figure is the mean (%) and the second the class ranking.

The unadjusted and adjusted (by ANCOVA) mean post-test scores of the three I.M.I. classes were compared. There was no significant difference in the distribution of the three classes; unadjusted $F(2,87) = .32$, $p > .05$, adjusted with P.A.T. $F(2,86) = .39$, $p > .05$, adjusted with Pre-test $F(2,86) = .80$, $p > .05$. These results suggest that the three I.M.I. classes performed similarly on the post-test although there were differences in the procedures used within each class.

One feature of this experiment was the differing exposure to the course content received by the students. The traditionally taught class was teacher-paced so all students were exposed to the total course content (from which the post-test was constructed). However, in the

I.M.I. classes student progress was a function of each individual student so that only a few students were exposed to the total content of the course.

This exposure to course content may be thought of as the 'opportunity to learn' (O.T.L.). This measure was defined by the possible score on questions corresponding to units the student had completed as a percentage of the total possible. The mean percentages of the O.T.L. for the three I.M.I. classes are given in Table 11. There was no significant difference in the distribution of the three classes on this measure of O.T.L., $F(2,88) = .30$, $p > .05$.

TABLE 11

The means and standard deviations of the opportunity to learn measure for the three I.M.I. classes.

Class	N	$\bar{X}$(%)	S.D.
A	29	74.2	21.8
B	32	76.5	20.5
D	30	72.2	24.7

Class A was mastery with proctors, B no mastery, D mastery.

In view of this differing O.T.L. for the students in the traditional class and the I.M.I. classes the post-test data was re-analysed. The traditionally taught classes results were left to stand as all students had been exposed to the course content. The score for each student in the I.M.I. classes was adjusted for exposure by marking only those questions which corresponded to units which they had completed, while questions corresponding to units which the student had not completed were ignored. It should be noted that the questions in the post-test had very similar (but not identical) format and content to questions in the unit tests, so reasonable discrimination was possible between whether a question was level 1 or level 2.

Three factors may explain students success in answering items from units they had not completed. Firstly, they may have done similar work in previous years or in other areas of their curriculum (e.g. Science). Secondly, an item may have been at level two while the student had only completed level one but had successfully transferred his knowledge to the upper level. Thirdly, the question may have come from a unit which the student had started but not completed for various reasons (see pages of this thesis.) This non-completion meant that the unit was not counted in the O.T.L. measure.

The exposure adjusted score for individuals in the experimental classes were obtained by dividing the mark obtained by the possible mark for 'exposed' questions.

The mean exposure adjusted post-test scores are given in Table 12. There was no significant difference in the distributions for the four classes, $F(3,116) = .26$ $p > .05$. This result suggests that if the variable opportunity to learn is taken into account, there is no difference among the classes in achievement.

TABLE 12

The means and standard deviations of the exposure adjusted post-test score for the four classes.

Class	N	$\bar{X}$(%)	S.D.
A	29	67.3	17.6
B	32	70.0	15.2
C	30	67.0	21.7
D	29	66.4	14.1

For the three I.M.I. classes the correlations of exposure adjusted post-test, P.A.T., pre-test and O.T.L. scores are given in Table 13.

TABLE 13

Intercorrelations of P.A.T., pre-test, post-test exposure adjusted post-(expost) and opportunity to learn (O.T.L.) for the three I.M.I. classes.

	PAT	pre	post	expost	O.T.L.
Pat		.68	.68	.60	.61
Pre			.76	.73	.66
Post				.93	.90
Expost					.74

These correlations suggest that the pre-test p (83) = .73 is a slightly better predictor than P.A.T. p (88) = .60 for the exposure adjusted post-test score (as it is for the non-adjusted post-test).

The opportunity to learn is a measure of the progress through the programme, it correlated highly with the unadjusted post-test, p (88) = .90. This correlation indicates that the students who make most rapid progress through the individualised programme performed best on the unadjusted post-test.

The opportunity to learn also correlated highly with the exposure adjusted post-test, p (88) = .74. This indicates again that students who made most rapid progress through the programme performed best on the adjusted post-test. However, the lower correlation would seem to suggest that if the exposure adjusted score indicates mastery of the course content covered then mastery is less dependent than achievement on progress through the programme.

CHAPTER IV

DISCUSSION

1. Post-course achievement as a measure on instructional effectiveness

Using the post-course test achievement as an evaluative measure of the effectiveness for teaching methods research has at least one serious flaw. The post-test is an attempt to measure what was learnt. However, considering that the duration of the experiment was a full academic year, what was more likely being measured was the student's memory and general revising skills, both of which were probably independent of any mode of instruction. That students in the two mastery classes invariably scored below previously attained levels on items similar to those supposedly mastered earlier in the course may be interpreted as evidence supporting the above contention. Further, the high correlations between the pre-course measures and the post-test may also indicate that post-test achievement is a function of something other than the mode of instruction.

The unadjusted (for O.T.L.) post-test results of this experiment fail to support the earlier finding of the superiority of individualised instruction with this level of student in this setting (Coppen 1976).

The present findings may be interpreted as support for the objections to Coppen's (1976) earlier study. The objections concerned content, teacher variability, bias due to exposure of items similar to those in the post-test and the method of constructing the post-test, as were discussed on page 25 of this thesis.

The problem of quality of instruction has been only partially controlled for this study. The study guide for the I.M.I. classes was used verbatim in the traditional class lecture and the same worked examples and tests were used in both groups. However, questions relating to the types and number of student queries asked in both groups and to the relative student activities were not investigated. These would require an observational study.

If the reason for the superiority of the traditional class is greater exposure to course content (and this is not just a statistical artifact), then serious consideration needs to be given to individualised instruction, if student academic achievement is the desired outcome.

The similarities of the exposure adjusted means (either adjusted for the pre-course measures or not) and the high correlations between O.T.L. and the post-test scores for the three I.M.I. classes provide support for Walker and Schaffarzick's (1974) contention that the most important variable in classroom learning is the exposure to material and not the mode of instruction.

Students in all I.M.I. classes had to satisfy the teacher that they had completed the required exercises for a topic/unit before they were given the appropriate topic/unit test so there was a certain degree of implied mastery in all I.M.I. classes. It may be questioned whether the teacher acting as a 'gate-keeper' results in similar 'first time' test scores and if the requirement of reaching mastery on subsequent tests adds anything to student learning. Unfortunately no count of either the relative number of 'failed' tests or the initial test results for the classes were kept. The results of the I.M.I. class (B) which did not require student mastery compared to Class (D) which did (but did not have student proctors) would seem to suggest that any initial academic 'edge' resulting from the attainment of mastery was transient, non-existant or unimportant in relation to other student processes involved in responding to post-test items. The contention of transience of mastery is supported in that many students in the two mastery classes scored below 80% on the exposure adjusted post-test. Had mastery been retained it could be argued that students in the two mastery classes should score at least at their previous level on items they had 'mastered'.

However, the question of whether mastery was retained is difficult to answer unequivocally from the results of this study. A problem for this study is the definition of mastery. The level taken as defining mastery was obtaining 80% on a unit test. However, as items varied in their weighting as graded by the teacher, this mark of 80% is quite arbitary. For example, had the items been weighted differently a mark of 80% may have changed to say 60% non-mastery, or even to 90%. In addition to the problem of defining what constituted mastery, there is the problem that although a student had obtained mastery (80% or greater) on a unit test he may have failed to master just those items on the test which was chosen as representative of that unit for the criterion referenced test. Thus it would not be possible to test, in this instance at least, if the student had retained mastery. Unfortunately the unit tests were retained by the students and therefore it was not possible to investigate this hypothesis.

The above difficulties not withstanding, if mastery had had any effect then it would seem reasonable that the mastery classes would have had higher means than the non-mastery classes (both the non-mastery I.M.I. class and the traditionally taught class) on the post-test adjusted for exposure to course content.

But as students in all four groups obtained similar distributions of scores when adjusted for exposure to course content, it would seem reasonable to argue that exposure per-se was the important factor and not mastery.

For the two classes with the mastery component (A and D) the results seem to suggest that tutoring (A) has little effect on achievement of the class overall. The six student proctors in Class A were ranked first to fourth, ninth and thirteenth on the pre-test scores and first to fifth and seventh on the post-test scores. In three I.M.I. classes only five students completed the entire course at ALL levels three of these being student proctors in Class A plus one student from the other two classes. Being a tutor, therefore, could not be said to be detrimental to student achievement and it certainly reduced the teacher's working load.

This study sought to evaluate an individualised mathematics programme using student academic achievement as a measure of effectiveness.

One of the underlying assumptions of teaching methods research is that what is learnt (as measured on tests of student recall) is a function of how it is taught with a corollary of this being that there are some methods of teaching which are better than others. However it can be asked whether a difference in student achievement should be expected on the grounds of differing instruction ?

When over 80 years of teaching methods research is reviewed (Stephens 1967) the ambivalent results of Miller's and Schoen's reviews are found to be the norm rather than the exception, where the dependent variable is student academic achievement. Invariably it is found that in any review of a particular innovative method there are students for whom the innovation is superior (*ca* 25%), others where there was no difference (*ca* 60%) and those where the traditional approach was superior (*ca* 15%). In view of these results it would appear that a difference between groups exposed to different instructional methods should not be expected. Stephens interpreted these results of the 80 years of teaching research as evidence that human interaction of teachers and students in the classroom with respect to learning is basically independent of the mode of instruction. He claimed that teachers invariably accentuate the important aspects ofthe curriculum to be learnt (and later tested) and that they positively reinforce learning by approval. When using an end-of-year test to compare groups of students it is possible that what is being tested is not the effectiveness of the mode of instruction but more likely student memory and general ability. Further, there are the problems that the students are also being exposed to learning situations such as help from parents, siblings, friends, and other students which are independent of the mode of instruction.

Walker and Schaffarzick (1974) concluded from their review of curriculum innovations that the only variable so far identified in academic learning was exposure to content (or opportunity to learn). They argue in their paper that by the time the student reaches the stage of formal education he/she has already obtained the foundations of how to learn. These learning processes which include the ability to sift information, ask relevant questions, obtain help from other children and the motivation to learn combine to produce a level of achievement that is 'usually far greater than any additional increment that might be produced by any further refinement of curricula or improvement in teaching style or method or medium of instruction'.

Thus, if as Stephens and Walker and Schaffarzick have argued, the major set of variables in learning reside somewhere in the learner (and hence are most probably amenable to scientific investigation) and if in a comparison study both groups are exposed to the same curricula then any differences of academic achievement between the two groups would not be expected.

If it is accepted that the major variables for academic learning reside somewhere in the learner, then it is unlikely that the contingencies related to the acquisition of such behaviour can be manipulated by the teacher. So does the area of contingency management have any relevance for the acquisition of academic learning ? From the previous arguments it would seem that the answer is no. Thus attention needs to be focused on stimulus control rather than the favoured area of contingency management.

Most of the studies in the area of academic behaviour reported during the last ten years in journals such as the Journal of Applied Behaviour Analysis concentrate on contingency mangement of skills already acquired, there being a dearth of studies involving the acquisition of new learning.

2. Are there 'best methods of learning'?

The corollary to the assumption that what is learnt is a function of how it is taught is that there are 'best methods of teaching'. Although common sense would suggest that there ought to be 'best methods of teaching' there is a view that in certain circumstances, such as those involving open systems, there will be no 'best way' (Katz and Kahn, 1966).

The general systems approach (von Bertalanffy, 1968) to questions involving educational research seems to have been largely neglected by educational research writers. General Systems theory (von Bertalanffy, 1968) views systems as being basically of two types, closed and open. A closed system is self-contained and independent of the external environment. This type of system is characteristically found in the physical sciences. Open systems, however, are acutely dependent upon their external environment. Biological organisms and social organisations are examples of such systems.

A consequence of a system being classified as open, is the principle of equifinality. This means that an open system can reach the same final state from differing initial conditions and by a variety of pathways. In other words, the open system can reorganise its internal subsystems to cope with the differing inputs to reach the same final state. In contrast in a closed system, the same initial conditions must lead to the same final result (Katz and Kahn, 1966).

From the previous discussion of skills and knowledge that students bring to the classroom and the academic interaction with peers, siblings and parents both inside and outside of the classroom it would seem dubious to claim that the classroom is isolated from the external environment as regards academic learning. The 'cognitive' boundaries of the classrooms are permeable to information and skill inputs particularly over long time spans. It would therefore seem reasonable to consider the classroom as an open system with regard to academic achievement. So, from this analysis it would appear that any attempt to find the one 'best way' of teaching is destined to fail !

After consideration of the arguments about the similarities of classroom interactions, whether the major set of variables in learning reside somewhere inside the learner (Stephens,1967, Walker and Schaffarzick, 1974) and if the classroom can be classified as an open system and hence whether a 'best' method of teaching should be expected, then not only does the question need to be asked whether simple outcome measures such as student achievement are the most appropriate but if they are even appropriate to judge a programme's worth at all. It would appear that no difference should be expected on such simple outcome measures as student achievement, which are the cornerstone of the conventional evaluation strategy, and further, that this strategy fails to consider many other facets of programmes such as teacher and student behaviour both in and out of the classroom, teacher and student attitudes to their new roles and the readability of the materials. It must be asked if there are alternative approaches to evaluation other than simple comparisons.

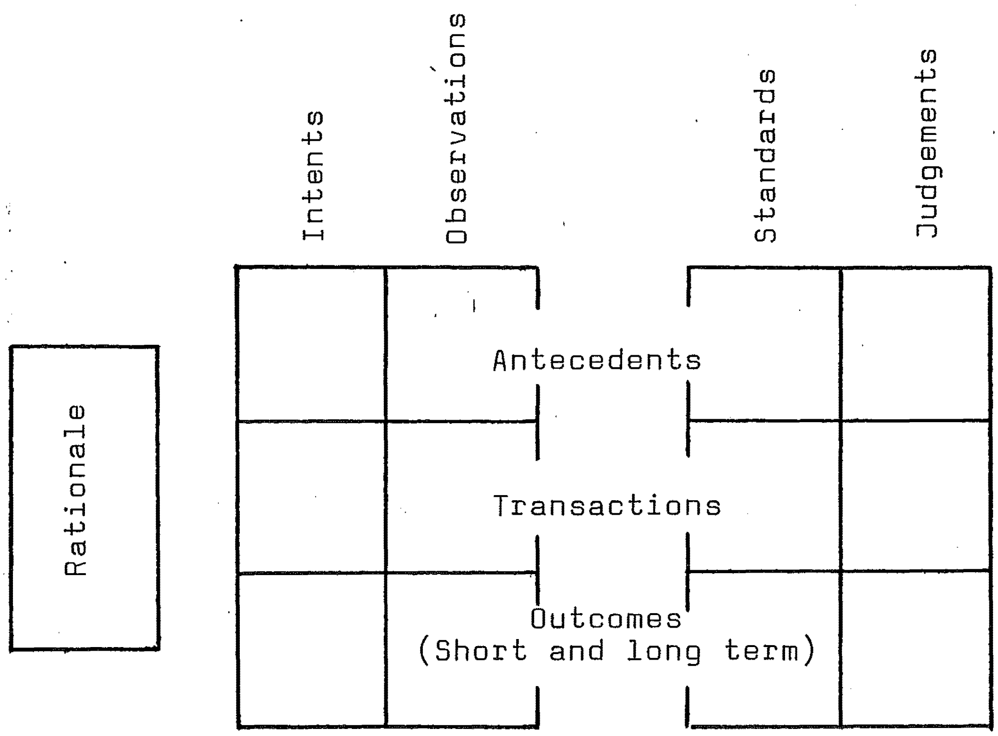


Figure 4
Stake's evaluation model

3. An alternative approach to the evaluation of an individualised mathematics course.

Stake (1967), responding to criticisms of large scale comparative studies (Cronback 1963), has developed a framework which explicitly investigates a number of variables in a single setting. This model shown in Figure 4 examines facets of programmes other than simple outcome measures.

When applying such a framework to the evaluation of an I.M.I. programme the rationale would be the designer's notions of long term objectives, behaviour of students and teachers and would be guided by at least an implicit idea of how the programme works. If these ideas are then made explicit they are open to scrutiny by interested parties and can be checked against empirical knowledge of the particular IQ, and/or age groups which are to use the programme. By making the theory explicit a framework is produced with which to guide modifications that are required when observations vary markedly from intents, or to rephrase it, if we think we know how the system works then we have some initial idea of how to fix it when it does not reach its objectives. A further benefit from explicating a theory is that information collected which is incongruent with the theory can lead to modification of the theory thereby possibly increasing our understanding the processes involved in individualised mathematics programmes.

Antecedents in Figure 4 refer to such variables as: length and type of teacher training, size of class, type of furniture, provision for I.M.I. Materials, student age and abilities.

Transactions refer to descriptions of the type of teacher and student behaviours which are a consequence of how the Individualised Mathematics programme is supposed to work (it would be anticipated that some idea of both type and quantity of behaviours would be specified). For the teacher these behaviours would include such activities as preparation and marking outside the classroom as well as such behaviours inside the classroom as distribution of tests, study guides and individual tutoring of students. For the student the type of behaviours would include on and off-task, as well as time spent waiting for tutorial help from the teacher in the classroom and out-of class the time spent on homework.

Outcome refers to the type of student achievement variables that would be measured at the end of the period of instruction (most probably at the end of the academic year) as well as sampling effective outcomes of both teacher and students (for example did the teachers like their new role, what did the students like most and least about the course, are the students' attitudes towards learning enhanced or deteriorating ?)

Once the course has been set up physically, the second phase of research, collating the observational data, can take place. For antecedents this would be to record the actual size of the class, the amount of teacher training, the ages and abilities of the students, as well as any unplanned for characteristics which may become obvious (e.g. internal arrangement of furniture).

The observation of the transactions would involve such things as teachers and students keeping a log of their activities outside the classroom as well as direct classroom observations of both teacher and student behaviours. By recording the type and number of questions asked during class time by the students, as well as actual response on the programme tests, information could be collected as to the difficulty of the content, so allowing the material to be altered where appropriate.

Related to the content difficulty is the readability of the instructional materials. By conducting suitable tests (eg the 'cloze' procedure, Rankin and Culhane, 1969) on material with the particular level of students concerned, the readability or otherwise of the materials may be ascertained. In this phase of the evaluation it would be possible to use participant observation (Smith and Geoffry,1968) to investigate what the students were comprehending and how they were using the instructional materials. How the writer of the materials and the classroom teacher think the students comprehend and use the material may be completely at variance with reality, so information about these student behaviours would be invaluable in producing more effective materials. The participant observer could also investigate the feedback and control mechanisms that students themselves employ when using that material, in particular if and how the various levels of students know that they understand the content.

The outcome observations would involve such activities as measuring the various students' achievement levels, the degree of retention of the course covered, their attitude to mathematics and this type of course as well as the problems they perceive with the programme content and format, during and at the end of the academic session. Interviews and questionaires could also be used to obtain information about how the teachers viewed their new roles and problems they perceived in the programme content and format.

The third and fourth phases of the evaluation fall into the areas of standards and judgements and basically involve decisions about the suitability of the level of content and design, as well as the importance of (or lack of) the degree of congruence between the intents and observations. The course content could be judged both by practising teachers and such interested parties as subject matter specialists and/or teachers college lecturers. For example, are the study guides detailed enough to convey the content to the designated students, did they correspond to the stated goals of the programme and were the test items in fact testing the goals enunciated ?

In the area of transactions, different audiences will arrive at differing conclusions when considering such variables as amount of teacher preparation time, amount of time spent with individual students as well as the way the students spent <u>their</u> time. Obviously interpretation of the relative importance of the various measures will depend on the value systems of the various audiences. At this stage the overall programme can be altered to mirror a consensus, particularly where the congruence between intended and observed outcomes is considered poor. To many this lack of a clear decision with respect to measuring merit of the individualised mathematics programme will seem to be a decisive characteristic against using this type of evaluation strategy, but this is the price which must be paid when we move from a single criteria for judging a programme to the larger real life picture.

Rather than stop after a single evaluation it is envisaged that the cycle described would be repeated with the idea of continually improving both the actual methods and materials as well as gaining insight into how the programme works.

A basic problem with this alternative evaluation strategy is the cost both in time and money as this strategy relies so heavily on observational techniques for collection of data and requires trained observers who are not readily available in New Zealand. However, if we wish to evaluate and improve individualised programmes then we must be prepared to pay for it.

REFERENCES

BERTALANFFY, L. von General Systems Theory. New York, George Braziller, 1968.

CARROLL, J.B. A Model of school learning. Teachers College Record, 1963, 64, 723-733.

COPPEN, M.J. Development and evaluation of an individualised mastery programme in mathematics for low-ability secondary school students. New Zealand Journal of Educational Studies, 1976 11, 132-142.

CLOWARD, R.D. Studies in tutoring. Journal of Experimental Education, 1976, 36, 14-25.

CROWBACH,L.J. Course improvement through Evaluation. Teachers' College Record, 1963, 64, 672-683.

CROSBY, J. cited in R.L. Miller, Individualised instruction in mathematics: a review of research. The Mathematics Teacher, 1976, 69, 345-351.

DINEEN, J.P. CLARK, H.B. and RISLEY T.R. Peer tutoring among elementary students: educational benefits to the tutor. Journal of Applied Behaviour Analysis, 1977, 10, 231-238.

EVANS, S.H. and ANASTASIO, E.J. Misuse of analysis of covariance when treatment effect and covariate are confounded. Psychological Bulletin, 1968, 69, 225-234.

HAMBLIN, R.L., HATHAWAY, C. and WODARSKI, J. Group contingencies, peer tutoring and accelerated academic achievement. In E.A. Ramp, and B.L. Hopkins (Eds), A new direction for education: Behaviour analysis - 1971, vol. 1. Lawrence, University of Kansas, 1971, Pp. 41-53.

HARRIS, V.W. and SHERMAN, J.A. Effects of peer tutoring and consequences on the math performance of elementary classroom students. Journal of Applied Behaviour Analysis, 1973, 6, 587-597.

JOHN, D.L. Paper presented at the meeting of the principals advisory committee, P.P.T.A. Christchurch, 1977.

JOHNSON, M and BAILEY, J.S. Cross-age tutoring: fifth graders as arithmetic tutors for kindergarten children. Journal of Applied Behaviour Analysis, 1974, 7, 223-232.

KATZ, D and KAHN, R.L. The Social Psychology of Organisations. New York, Wiley, 1966.

MILLER, R.L. Individualised instruction in mathematics: a review of research. The Mathematics Teacher, 1976, 69, 345-351.

NEALE, J.M. and LIEBERT, R.M. Science and behaviour an introduction to methods of research. Englewood Cliffs, Prentice-Hall, 1973.

NIGHTINGALE, D.A., CORNWALL, R.M., GRAY, M.J., KIBBLEWHITE, D.F. and SMITH, S.J. Mathematics: a study in Pattern 1. Auckland, Jacaranda, 1972.

RANKIN, E.F. and CULHANE, J.W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.

ROSENTHAL, R. Experimenter bias in behavioural research. New York, Appleton-Century-Crofts, 1966.

SCHOEN, H.L. Self paced mathematics instruction, how effective has it been in secondary and post-secondary schools ? The Mathematics Teacher, 1976, 69, 352-357.

SCRIVEN, M. The methodology of evaluation. In B.R. Worthen and J.R. Sanders (Eds), Educational evaluation: theory and practice, Belmont (California), Wadsworth, 1973.

SCRIVEN, M. Problems and prospects for individualisation. In H. Talmage (Ed), Systems of individualised instruction, Berkeley (California), McCutchan, 1975.

SMITH, L.M. and GEOFFREY, W. The complexities of an urban classroom, New York, Holt, Rinehart and Winston, 1968.

STAKE, R.E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

STEPHENS, J.M. The process of schooling, Toronto, Holt, Rinehart and Winston, 1967.

WALKER, D.F. and SCHAFFARZICK , J. Comparing Curricula. Review of Educational Research, 1974, 44, 83-111.

WHITE, V.T. A profile of individualised instruction. The Mathematics Teacher, 1972, 65, 394, 448-453.

APPENDIX

THIRD FORM MATHEMATICS $L_{1,2}$ TIME 2HR

NAME ____________ CLASS ______

SECTION A

NO WORKING NEED BE SHOWN; BUT IF YOU DO WORKING, DO IT NEATLY ON PAD PAPER AND HAND IT IN.

1 (a) FROM THE FOLLOWING REASONS INDICATE WHY THE FOLLOWING STATEMENTS ARE TRUE FOR FIGURE 1

[LIST OF REASONS: ADJACENT ANGLES ON A LINE, VERTICALLY OPPOSITE ANGLES, ANGLES AT A POINT.]

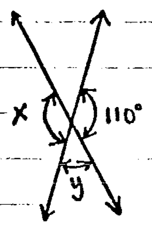


FIGURE 1

REASON

$x = 110°$ ____________

$y + 110° = 180°$ ____________

(b) FROM THE LIST AVAILABLE CHOOSE THE MOST APPROPRIATE NAME FOR THE ANGLE [ACUTE, RIGHT, OBTUSE, STRAIGHT, REFLEX

(i) $A = 90°$ A IS ________

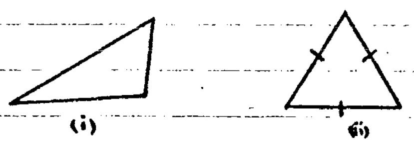
(ii) $B = 200°$ B IS ________

[4 MARKS]

2 (a) THE SUM OF THE INTERIOR ANGLES OF A TRIANGLE IS ______°

(b) IF ONE INTERIOR ANGLE OF AN ISOSCELES TRIANGLE IS 100° THEN WHAT ARE THE SIZES OF THE OTHER TWO ANGLES? ____° AND ____°

(c) GIVE THE MOST APPROPRIATE NAMES FOR EACH TRIANGLE SHOWN IN FIGURE 2 (HELPFUL NAMES: SCALENE, ISOSCELES, EQUILATERAL, ACUTE, RIGHT, OBTUSE



NAMES

(i) ____________

(ii) ____________

FIGURE 2

[4 MARKS]

3(a) GIVE REASONS WHY THE FOLLOWING ARE TRUE STATEMENTS FOR FIGURE 3 [HELPFUL NAMES: ALTERNATE, CORRESPONDING, SUM INTERIOR ANGLE TRIANGLE, EXTERIOR ANGLE SUM INTERIOR OPPOSITE ANGLE]

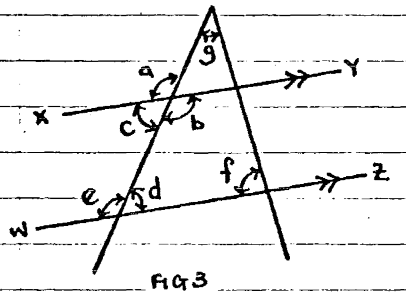


FIG 3

	REASON
$a = e$	
$c = d$	
$b = a$	
$a + c = 180°$	
$e = g + f$	
$g + d + f = 180°$	

NOTE $\overleftrightarrow{XY} \parallel \overleftrightarrow{WZ}$

(b)

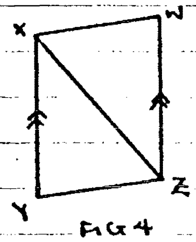


FIG 4

ON FIG. 4 NAME A PAIR OF EQUAL ALTERNATE ANGLES ______ AND ______

[7 MARKS]

4(a) FOR THE POLYGONS SHOWN IN FIG 5 CHOOSE THE MOST APPROPRIATE NAME FROM THE FOLLOWING LIST: TRIANGLE, QUADRILATERAL PENTAGON, HEXAGON, OCTAGON

NOTE UNLESS SIDES AND OR ANGLES ARE INDICATED TO BE EQUAL DO NOT ASSUME THAT THEY ARE!

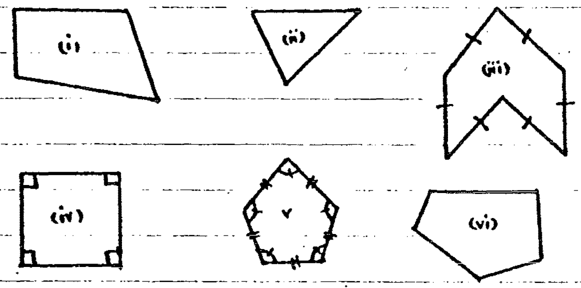


FIGURE 5

NAMES

(i) ______

(ii) ______

(iii) ______

(iv) ______

(v) ______

(vi) ______

(b) LIST ALL THOSE POLYGONS IN FIGURE 5 WHICH ARE:

(i) REGULAR ______

(ii) RE-ENTRANT ______

(c) THE SUM OF THE INTERIOR ANGLES OF A PENTAGON IS: ______°

(d) THE SUM OF THE EXTERIOR ANGLES OF A TRIANGLES IS: ______°

[7 MARKS]

5(a)

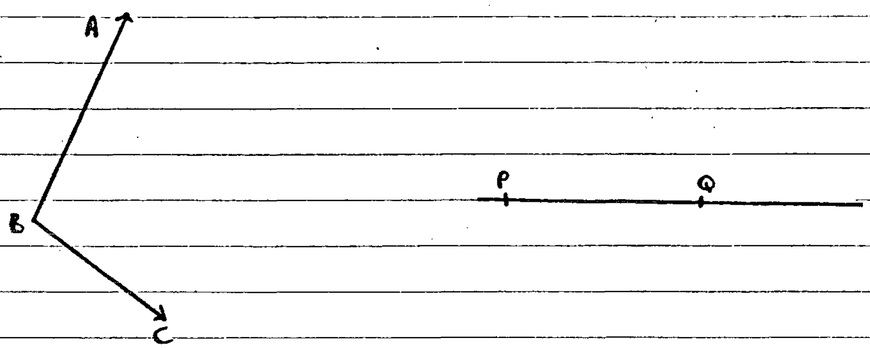


FIGURE 6(a)

USE A COMPASS AND RULER ONLY FOR THIS QUESTION (5a) LEAVE IN ALL CONSTRUCTION LINES AND ARCS.

(i) ON FIGURE 6 CONSTRUCT THE ANGLE BISECTOR OF $A\hat{B}C$

(ii) ON FIGURE 6 CONSTRUCT THE BISECTOR OF $\overline{PQ}$, SO LABEL THE MIDPOINT OF $\overline{PQ}$, M

FIGURE 6b

(b) ON FIGURE 6(b) (i) USING A COMPASS CAREFULLY CONSTRUCT THE FOLLOWING A CIRCLE, CENTRE B, WITH RADIUS 5 cm THEN A CIRCLE, CENTRE C, WITH RADIUS 6 cm

(ii) LABEL THE TWO POINTS WHERE THE CIRCLES CUT P AND Q

(iii) USING YOUR RULER DRAW IN THE FOLLOWING LINE SEGMENTS $\overline{PB}$ THEN $\overline{PC}$

(iv) USING A PROTRACTOR MEASURE ANGLE $B\widehat{P}C$, ______ °

(v) AT POINT C DRAW IN AN EXTERIOR ANGLE OF THE TRIANGLE PBC. MARK THE ANGLE CLEARLY WITH AN X

[8 MARKS]

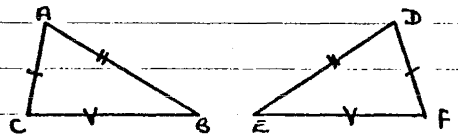
6(a) IF $\Delta HIJ \cong \Delta MON$ THEN COMPLETE (i) $\overline{HI} \cong$ ____

(ii) $\widehat{MNO} \cong$ ____

(b) FOR EACH PAIR OF TRIANGLES IN FIGURE 7 INDICATE WHETHER OR NOT THEY ARE CONGRUENT BY PLACING A CIRCLE AROUND THE Y OR N

IF THE PAIR ARE CONGRUENT, THEN GIVE THE CORRECT REASON [S.S.S, ETC] AND COMPLETE THE CONGRUENCY EXPRESSION

(i)

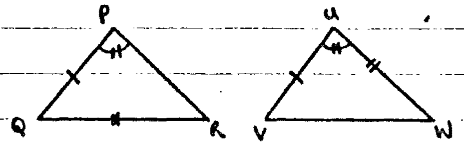


CONGRUENT ? Y OR N

REASON ____

$\Delta ABC \cong \Delta$ ____

(ii)



CONGRUENT ? Y OR N

REASON ____

$\Delta PQR \cong \Delta$ ____

[6 MARKS]

FIGURE 7

7(a) WHAT IS THE PROBABILITY OF A CERTAIN EVENT ? Pr(CERTAIN) = ____

(b) IF A SINGLE SIX SIDED DIE IS ROLLED WHAT IS THE PROBABILITY THAT THE RESULT WILL BE A THREE ? Pr (THREE) = ____

(c) A BOX CONTAINS THREE RED BALLS AND FIVE WHITE BALLS IF A BALL IS DRAWN OUT OF THE BOX WHAT IS THE PROBABILITY THAT IT WILL BE A RED ONE ? Pr (RED) = ____

(d) A SINGLE COIN IS TOSSED 180 TIMES, WHAT IS THE EXPECTED NUMBER OF HEADS UPPERMOST ? EXPECTED Nº OF HEADS = ____

[4 MARKS]

8(a) ROSTER THE LETTERS OF THE WORD "ELEMENT" {.....................}

(b) IF $A = \{1,2,3\}$, $B = \{2,3,4,5\}$, $C = \{5\}$

(i) PLACE THE ELEMENTS (NUMBERS) IN THE CORRECT REGIONS IN THE VENN DIAGRAM IN FIGURE 8

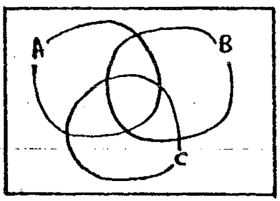


FIGURE 8

(ii) ROSTER THE ELEMENTS OF $A \cap B$ = {.....................}

(iii) ROSTER THE ELEMENTS OF $A \cup B$ = {.....................}

(iv) $n(A)$ IS ? $n(A) =$ ____

(v) $n(A \cap C)$ IS ? $n(A \cap C) =$ ____

(c) FROM THE VENN DIAGRAM IN FIGURE 9 ROSTER THE FOLLOWING SETS

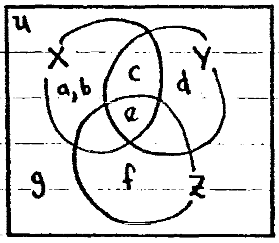


FIGURE 9

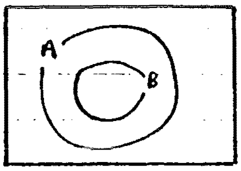
(i) U = {......................}

(ii) X = {......................}

(iii) X ∩ Y = {......................}

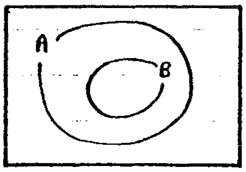
(iv) Y ∩ Z = {......................}

(d) ON THE VENN DIAGRAMS GIVEN IN FIGURE 10 SHADE THE AREAS DESCRIBED



A ∩ B

FIG 10



A ∪ B

(e) IDENTIFY THE NUMBER SYSTEMS GIVEN FROM N, W, I, Q, R

{ ... -1, 0, 1, 2, 3 ... } ______

{ 1, 2, 3, ... } ______

{ 0, 1, 2, ... } ______

[15 MARKS]

9(a) SIMPLIFY THE FOLLOWING, GIVING YOUR ANSWERS IN THE <u>SIMPLEST</u> POSSIBLE FORM

(i) $3(14-6)$ (i) ______

(ii) $15-3+4$ (ii) ______

(iii) $4+6\times3$ (iii) ______

(iv) $4+4(4-4)$ (iv) ______

(v) $^{-}2-{}^{-}2$ (v) ______

(b) EXPAND THE FOLLOWING, THAT IS WRITE THE PRODUCT AS A SUM

$3(2-A)$ (b) ______

(c)(i) WRITE 6^2 AS A MORE COMMON NUMBER (i) ______

(ii) WRITE 16 IN INDEX FORM WITH THE SIMPLEST POSSIBLE BASE (ii) ______

(d) IF $a=0$, $b=1$, $c=2$ THEN EVALUATE THE FOLLOWING

(i) $3a$ (i) ______

(ii) $b-a\times c$ (ii) ______

(iii) $3c^2$ (iii) ______

(iv) $(2b)^2$ (iv) ______

[12 MARKS]

10 (a) COMPLETE THE FOLLOWING SEQUENCE OF ORDERED PAIRS

(4,2), (5,3), (6, ____), ... (19, ____), ... (n, _____)

(b) (i) ON FIGURE 11 PLOT THE POINT (3,1) AND LABEL IT A

ALSO PLOT THE POINT (-1,4) AND LABEL IT B

(ii) WHAT ARE THE COORDINATES OF C IN FIGURE 11 C = (__, __)

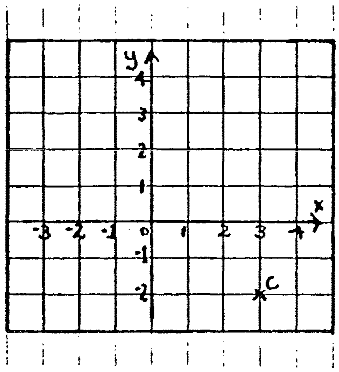


FIGURE 11

(c) A RELATION IS GIVEN BY THE SET OF ORDERED PAIRS

R = {(1,2), (2,4), (3,6)}

(i) ROSTER THE DOMAIN OF R {.............}

(ii) ROSTER THE RANGE OF R {.............}

[8 MARKS]

11. SIMPLIFY THE FOLLOWING

(a) (i) $\frac{4}{5} - \frac{1}{5}$ (ii) $\frac{1}{2} \times \frac{1}{3}$ (i) ____ (ii) ____

(b) (i) 14.6 − 3.45 ______

(ii) 5.06 + 7.94 ______

(iii) 3.2 × 10 ______

(iv) 23.6 ÷ 100 ______

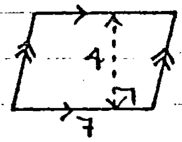
(v) 1000 × 4.01 ______

[7 MARKS]

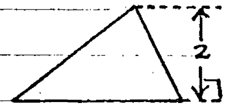
12 (a) CALCULATE THE AREAS OF THE FOLLOWING FIGURES

(ALL MEASUREMENTS ARE IN CENTIMETRES c.m.)

AREAS



(i)



(ii)

(i) ______

(ii) ______

(b) (i) CALCULATE THE VOLUME OF A CUBE OF SIDE 2cm V= ______

(ii) CALCULATE THE VOLUME OF THE RECTANGULAR BOX SHOWN IN FIGURE 13 (ALL MEASUREMENTS ARE IN CENTIMETRES)

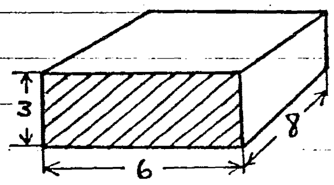


FIGURE 13

VOLUME = ______

[6 MARKS]

13(a) WRITE $3\times10^3 + 2\times10 + 6$ AS A MORE USUAL NUMBER ______

(b) (i) THE NUMBER OF MILLILITRES IN A LITRE IS? b(i) ______

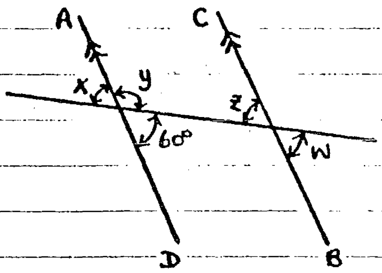
(ii) EXPRESS 4 KILOMETRES 3 METRES IN KILOMETRES (ii) ______

[3 MARKS]

SECTION B

FULL SETTING OUT IS REQUIRED FOR THESE QUESTIONS BE VERY SURE THAT YOU WORK CAN BE FOLLOWED

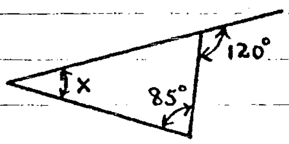
1(a) FIND THE SIZES OF THE UNKNOWN ANGLES IN FIGURE 14 FULL REASONS MUST BE GIVEN (SEE SECTION A Q1,3 FOR HELPFUL NAMES)



NOTE $\overline{AD} \parallel \overline{CB}$

FIG 14

(b) FIND THE SIZE OF X IN FIGURE 15, FULL REASONS MUST BE GIVEN (SEE SECTION A Q 3 FOR HELPFUL NAMES)



[11 MARKS]

2 SHOW HOW TO CALCULATE THE SIZE OF AN INTERIOR ANGLE OF A REGULAR NINE-SIDED POLYGON

AN INTERIOR ANGLE = ______°

[3 MARKS]

3 THE FOLLOWING MARKS WERE OBTAINED BY A SMALL GROUP OF STUDENTS ON A TEST 1, 0, 6, 3, 5, 3, 4, 2, 1

SHOW HOW TO FIND THE MODE, MEDIAN, MEAN

MODE = ______

MEDIAN = ______

MEAN = ______

IF THE PASS MARK WAS 4, HOW MANY STUDENTS PASSED? ______

4 (a) JOHN SCORED 10 RUNS AND THE TEAMS TOTAL WAS 50

EXPRESS JOHN SCORE AS A PERCENTAGE OF THE TEAMS TOTAL

JOHNS SCORE = ______ %

(b) A MAN EARNS \$3000. IF THE TAX DEPARTMENT WANT 20% OF HIS WAGES IN TAX, HOW MUCH DOES HE PAY

TAX = \$ ______

5(a) FOR $2X + 1 = 7$, $X \in \{2, 3, 4\}$

THEN ROSTER (i) TRUTH SET, T, (ii) FALSE SET T'

T = {........}

T' = {........}

(b) SHOW HOW TO SOLVE (DO NOT DO MENTALLY!)

(i) $X - {}^{+}6 = {}^{-}5$ 　　$X =$ ______

(ii) $Y - {}^{-}3 = {}^{-}4$ 　　$Y =$ ______